

第二章 随机变量及其分布

随机变量

离散型随机变量

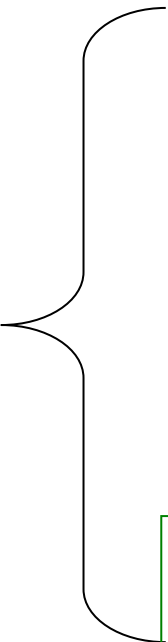
分布函数

连续型随机变量

随机变量的函数

2.1 随机变量

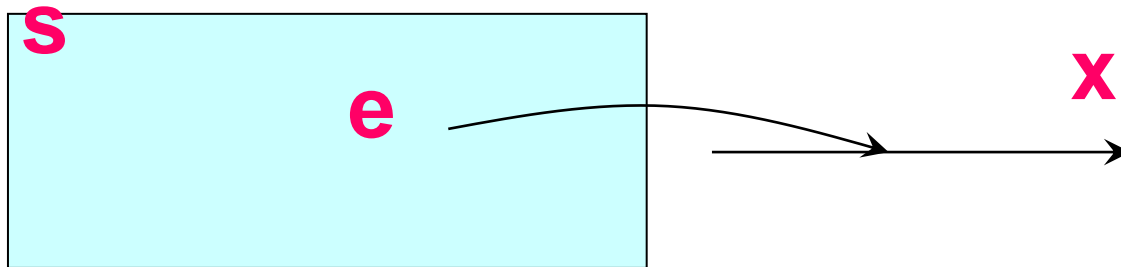
常见的两类试验结果：



示数的——降雨量；
候车人数；
发生交通事故的次数...

示性的——明天天气（晴，云...）；
化验结果（阳性，阴性）...

中心问题：将试验结果数量化



$X=X(e)$ ——为 S 上的实值单值函数

定义： 设随机试验的样本空间为 $S = \{e\}$ ，若 $X = X(e)$ 为定义在样本空间 S 上的实值单值函数，则称 $X = X(e)$ 为随机变量。

- 一般采用大写英文字母 X, Y, Z 来表示随机变量
- 引入随机变量的目的是用来描述随机现象

一般地, 若 I 是一个实数集合, 则
 $\{X \in I\}$ 为事件 $\{e: X(e) \in I\}$

常见的两类随机变量  离散型的
连续型的

例1.1 掷硬币**3**次，出现正面的次数记为 **X** .

样本点	TTT	TTH	THT	HTT	HHT	HTH	THH	HHH
X 的值	0	1	1	1	2	2	2	3

$$P\{X = 0\} = P\{TTT\} = 1/8$$

$$P\{X = 1\} = P\{TTH, THT, HTT\} = 3/8$$

$$P\{X \leq 1\} = P\{X = 0\} + P\{X = 1\} = 1/2$$

X	0	1	2	3
p	1/8	3/8	3/8	1/8

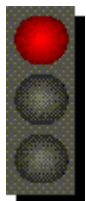
2.2 离散型随机变量及其分布

定义：取值至多可数的随机变量为离散型的随机变量。概率分布律为

X	x_1	x_2	\cdots	x_i	\cdots
p	p_1	p_2	\cdots	p_i	\cdots

概率分布律性质： $p_i \geq 0, \sum_{i=1}^{\infty} p_i = 1$

概率分布律 { 写出所有可能取值;
写出每个取值相应的概率.



例2.1 某人骑自行车从学校到火车站，一路上要经过3个独立的交通灯，设各灯工作独立，且设各灯为红灯的概率为 p ， $0 < p < 1$ ，以 X 表示首次停车时所通过的交通灯数，求 X 的概率分布律。

解：设 $A_i = \{\text{第}i\text{个灯为红灯}\}$ ，则 $P(A_i) = p$ ， $i=1, 2, 3$ 且 A_1, A_2, A_3 相互独立。

$$P(X = 0) = P(A_1) = p ;$$

$$P(X = 1) = P(\bar{A}_1 A_2) = (1 - p)p ;$$

$$P(X = 2) = P(\bar{A}_1 \bar{A}_2 A_3) = (1 - p)^2 p ;$$

$$P(X = 3) = P(\bar{A}_1 \bar{A}_2 \bar{A}_3) = (1 - p)^3 ;$$

X	0	1	2	3
p	p	$p(1-p)$	$(1-p)^2 p$	$(1-p)^3$

例2.2 若随机变量 X 的概率分布律为

$$P(X = k) = \frac{c\lambda^k}{k!}, \quad k = 0, 1, 2, \dots, \quad \lambda > 0$$

求常数 c .

解:

$$1 = \sum_{k=0}^{+\infty} P\{X = k\}$$

$$= c \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} = ce^{\lambda}$$

$$\Rightarrow c = e^{-\lambda}$$

几个重要的离散型随机变量分布

一、0-1分布

若 X 的分布律为:

X	0	1
p	q	p

随机变量只可能
取**0**、**1** 两个值

$$(p+q=1, p>0, q>0)$$

则称 X 服从**参数为 p 的0-1分布**，或两点分布.

记为 $X \sim 0-1(p)$ 或 $B(1, p)$

0-1 (p) 分布的分布律还可以写为

$$P(X = k) = p^k (1 - p)^{1-k}, \quad k = 0, 1.$$

对于一个随机试验，如果它的样本空间只包含两个元素，即 $S = \{e_1, e_2\}$ ，我们总能在 S 上定义一个服从 **(0—1) 分布** 的随机变量。

$$X = X(e) = \begin{cases} 0, & \text{当 } e = e_1, \\ 1, & \text{当 } e = e_2. \end{cases}$$

来描述这个随机试验的结果。

检查产品的质量是否合格，对新生婴儿的性别进行登记，检验种子是否发芽以及前面多次讨论过的“抛硬币”试验都可以用 **(0—1) 分布** 的随机变量来描述。

一个随机试验,设 A 是一随机事件,且
 $P(A)=p, (0 < p < 1)$.若仅考虑事件 A 发生与否,
定义一个服从参数为 p 的0-1分布的随机变
量:

$$X = \begin{cases} 1, & \text{若} A \text{发生,} \\ 0, & \text{若} A \text{不发生 (即} \bar{A} \text{发生).} \end{cases}$$

来描述这个随机试验的结果。

只有两个可能结果的试验,称为贝努利试验。

二、二项分布

n 重贝努利试验：设试验 E 只有两个可能的结果： A 与 \bar{A} ， $P(A)=p$ ， $0 < p < 1$ ，将 E **独立地重复**进行 n 次，则称这一串重复的独立试验为 **n 重贝努利试验**。

即每次试验结果
互不影响

在相同条件下
重复进行

- 独立重复地抛 n 次硬币，每次只有两个可能的结果：正面，反面，

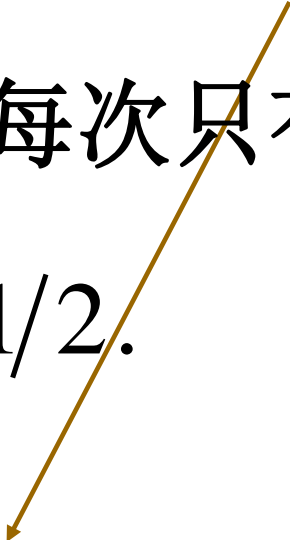
$$P(\text{出现正面}) = 1/2.$$

- 将一颗骰子抛 n 次，设 $A=\{\text{得到1点}\}$ ，则每次试验只有两个结果： A, \bar{A} ,

$$P(A) = 1/6.$$

从52张牌中有放回地取 n 次，设 $A=\{\text{取到红牌}\}$ ，则每次只有两个结果： A, \bar{A} ,

$$P(A) = 1/2.$$



如果是不放回抽样呢？
不是独立试验！

设 A 在 n 重贝努利试验中发生 X 次, 则

$$P(X = k) = C_n^k p^k (1-p)^{n-k}, k = 0, 1, \dots, n$$

并称 X 服从参数为 n, p 的二项分布, 记

$$X \sim B(n, p)$$

$$\text{注: } 1 = (p + q)^n = \sum_{k=0}^n C_n^k p^k q^{n-k} \quad \text{其中 } q = 1 - p$$

推导：以 $n=3$ 为例, 设 $A_i = \{ \text{第}i\text{次}A\text{发生} \}$

$$P(X = 0) = P(\bar{A}_1 \bar{A}_2 \bar{A}_3) = (1 - p)^3$$

$$P(X = 1) = P(A_1 \bar{A}_2 \bar{A}_3 \cup \bar{A}_1 A_2 \bar{A}_3 \cup \bar{A}_1 \bar{A}_2 A_3) = C_3^1 p^1 (1 - p)^{3-1}$$

$$P(X = 2) = P(A_1 A_2 \bar{A}_3 \cup A_1 \bar{A}_2 A_3 \cup \bar{A}_1 A_2 A_3) = C_3^2 p^2 (1 - p)^{3-2}$$

$$P(X = 3) = P(A_1 A_2 A_3) = p^3$$

一般 $P(X = k) = C_n^k p^k (1 - p)^{n-k}, k = 0, 1, 2, \dots, n$

例2.3 有一大批产品，其验收方案如下：
先作第一次检验，从中任取10件，经检验无次品接受这批产品，次品数大于2拒收；否则作第二次检验，从中任取5件，仅当5件中无次品便接受这批产品，设产品的次品率为 p ．求这批产品能被接受的概率．

解：设 $A = \{\text{接受该批产品}\}$ 。设 X 为第一次抽得的次品数， Y 为第2次抽得的次品数。

则 $X \sim B(10, p)$, $Y \sim B(5, p)$, 且 $\{X = i\}$ 与 $\{Y = j\}$ 独立。

$$\begin{aligned} P(A) &= P(X = 0) + P(1 \leq X \leq 2 \text{ 且 } Y = 0) \\ &= P(X = 0) + P(1 \leq X \leq 2) \cdot P(Y = 0) \\ &= P(X = 0) + (P(X = 1) + P(X = 2)) \cdot P(Y = 0) \\ &= (1 - p)^{10} + [10p(1 - p)^9 + 45p^2(1 - p)^8] \cdot (1 - p)^5 \end{aligned}$$

例2.4 设随机变量 $X \sim B(100, 0.05)$,
求 $P(X \leq 10)$ 和 $P(X = 10)$

$$\begin{aligned}\text{解: } P(X \leq 10) &= \sum_{k=0}^{10} P(X = k) \\ &= \sum_{k=0}^{10} C_{100}^k 0.05^k 0.95^{100-k}\end{aligned}$$

使用*Excel*表单:

在任一单元格中输入

“***=BINOM.DIST(10,100,0.05,TRUE)***”,
点“确定”后,在单元格中出现“**0.988528**”.
这里“*TRUE*”可用“1”代替.

计算 $P(X=10)$,

在任一单元格中输入

“***=BINOM.DIST(10,100,0.05,FALSE)***”,
点“确定”后,在单元格中出现“**0.016715884**”.
这里“*FALSE*”可用“0”代替.

三、泊松分布(Poisson分布)



若随机变量 X 的概率分布律为

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \dots, \quad \lambda > 0$$

称 X 服从参数为 λ 的泊松分布，记

$$X \sim P(\lambda)$$

例2.5 某公交站单位时间内候车人数

$$X \sim P(4.8),$$

求(1)随机观察1个单位时间，至少有3人候车的概率；

(2)随机独立观察5个单位时间，恰有4个单位时间至少有3人候车的概率.

解:(1) $P(X \geq 3) = 1 - P(X = 0) - P(X = 1) - P(X = 2)$

$$= 1 - e^{-4.8} \left(1 + 4.8 + \frac{4.8^2}{2!} \right) = 0.8580$$

(2) 设5个单位时间内有 Y 个单位时间是“至少有3人候车”，

则 $Y \sim B(5, p)$, 其中 $p = P(X \geq 3) = 0.8580$,

于是 $P(Y = 4) = C_5^4 p^4 (1 - p) = 0.7696$.

泊松定理:

二项分布与泊松分布有下面的近似结果

当 $n > 10, p < 0.1$ 时,

$$C_n^k p^k (1-p)^{n-k} \approx \frac{e^{-\lambda} \lambda^k}{k!}, \text{ 其中 } \lambda = np.$$

$$\begin{aligned}
 \text{事实上, } C_n^k p^k (1-p)^{n-k} &= \frac{n!}{k!(n-k)!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\
 &= \frac{\lambda^k}{k!} \frac{n(n-1)\dots(n-k+1)}{n^k} \left[\left(1 - \frac{\lambda}{n}\right)^{-n/\lambda} \right]^{-\lambda} / \left(1 - \frac{\lambda}{n}\right)^k \\
 &\approx \frac{e^{-\lambda} \lambda^k}{k!}
 \end{aligned}$$

因为当 n 充分大和适当的 λ 时,

$$\frac{n(n-1)\dots(n-k+1)}{n^k} \approx 1, \left(1 - \frac{\lambda}{n}\right)^k \approx 1, \left[\left(1 - \frac{\lambda}{n}\right)^{-n/\lambda} \right]^{-\lambda} \approx e^{-\lambda}$$

例2.6 某地区一个月內成年人患某种疾病的患病率为 $1/200$ ，设各人是否患病相互独立。若该地区一社区有1000个成年人，求某月內该社区至少有3人患病的概率。

解：设该社区1000人中有 X 个人患病，

$$\text{则 } X \sim B(1000, \frac{1}{200}),$$

$$\begin{aligned} P(X \geq 3) &= 1 - P(X = 0) - P(X = 1) - P(X = 2) \\ &= 0.8760; \end{aligned}$$

利用泊松分布进行近似计算，取 $\lambda = 5$,

$$P(X \geq 3) \approx 1 - \frac{e^{-5}}{0!} - \frac{5e^{-5}}{1!} - \frac{5^2 e^{-5}}{2!} = 0.8753.$$

泊松分布使用*Excel*表单:

在*Excel*的任一单元格输入

“***=POISSON.DIST(2,5,1)***”，回车，
就在单元格中出现 **“0.124652019”**。

$$P(X \geq 3) = 1 - P(X \leq 2) = 0.875347981.$$

四、超几何分布

若随机变量 X 的概率分布律为

$$P(X = k) = \frac{C_a^k C_b^{n-k}}{C_N^n}, k = l_1, l_1 + 1, \dots, l_2,$$

其中, $l_1 = \max(0, n - b)$, $l_2 = \min(a, n)$.

称 X 服从超几何分布.

例2.7 一袋中有 a 个白球, b 个红球, $a+b=N$, 从中不放回地取 n 个球, 设每次取到各球的概率相等, 以 X 表示取到的白球数, 则 X 服从超几何分布。

五、几何分布

若随机变量 X 的概率分布律为

$$P(X = k) = p(1 - p)^{k-1}, k = 1, 2, 3, \dots, \quad 0 < p < 1.$$

称 X 服从参数 p 的几何分布.

例2.8 从生产线上随机抽产品进行检测，设产品的次品率为 p ， $0 < p < 1$ ，若查到一只次品就得停机检修，设停机时已检测到 X 只产品，则 X 服从参数 p 的几何分布。

六、巴斯卡分布

若随机变量 X 的概率分布律为

$$P(X = k) = C_{k-1}^{r-1} p^r (1-p)^{k-r}, k = r, r+1, r+2, \dots,$$

其中 r 为正整数, $0 < p < 1$.

称 X 服从参数为 (r, p) 的巴斯卡分布.

例2.9 独立重复地进行试验，每次试验的结果为成功或失败，每次试验中成功的概率均为 p ， $0 < p < 1$ ，试验进行到出现 r 次成功为止，以 X 表示试验次数，则 X 服从参数为 (r, p) 的巴斯卡分布。

思考题：一盒中有2个红球4个白球，

(1) 从中取一球， X 表示取到的红球数；

(2) 采用不放回抽样取3球， Y 表示取到的红球数；

(3) 采用放回抽样取3球， Z 表示取到的红球数；

(4) 采用放回抽样取球，直到取到红球为止， U 表示取球次数；

(5) 采用放回抽样取球，直到取到3个红球为止， V 表示取球次数。

上述随机变量 X, Y, Z, U, V 的分布律是什么呢？

解答：(1) X 服从0—1分布， $P(X=1)=1/3$ ， $P(X=0)=2/3$ ；

(2) Y 服从超几何分布，

$$P(Y = k) = \frac{C_2^k C_4^{3-k}}{C_6^3}, k = 0, 1, 2;$$

(3) Z 服从二项分布 $B(3, 1/3)$ ，

$$P(Z = k) = C_3^k \frac{2^{3-k}}{3^3}, k = 0, 1, 2, 3;$$

(4) U 服从几何分布，

$$P(U = k) = \frac{2^{k-1}}{3^k}, k = 1, 2, 3, \dots$$

(5) V 服从巴斯卡分布，

$$P(V = k) = C_{k-1}^2 \frac{2^{k-3}}{3^k}, k = 3, 4, 5, \dots$$

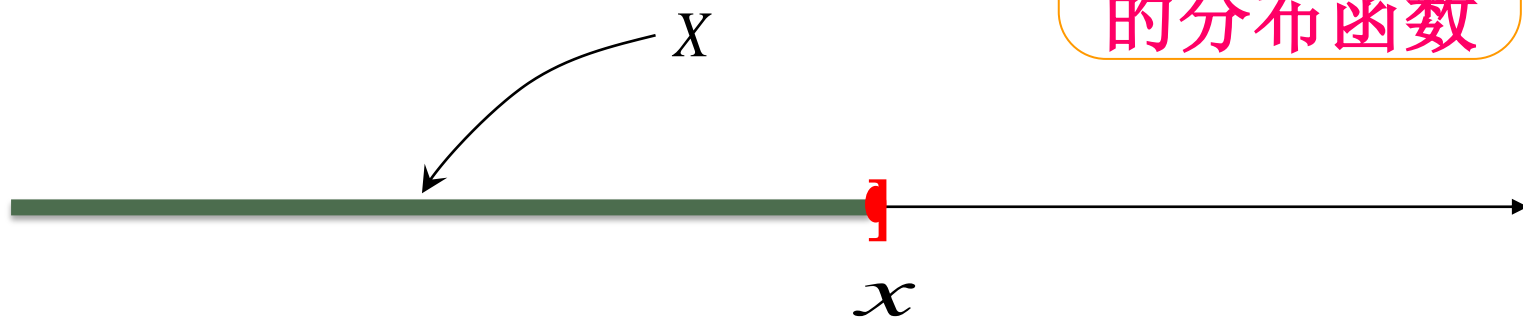
2.3 随机变量的分布函数

定义：随机变量 X ，若对任意实数 x ，函数

$$F(x) = P(X \leq x)$$

称为 X 的分布函数.

$F(x)$ 的几何意义：



分布函数的性质：

1) $0 \leq F(x) \leq 1$;

2) $F(x)$ 单调不减, 且 $F(-\infty) = 0$, $F(+\infty) = 1$

$$\because 0 \leq P(x_1 < X \leq x_2) = F(x_2) - F(x_1)$$

3) $F(x)$ 右连续, 即 $F(x+0) = F(x)$;

4) $F(x) - F(x-0) = P(X = x)$.

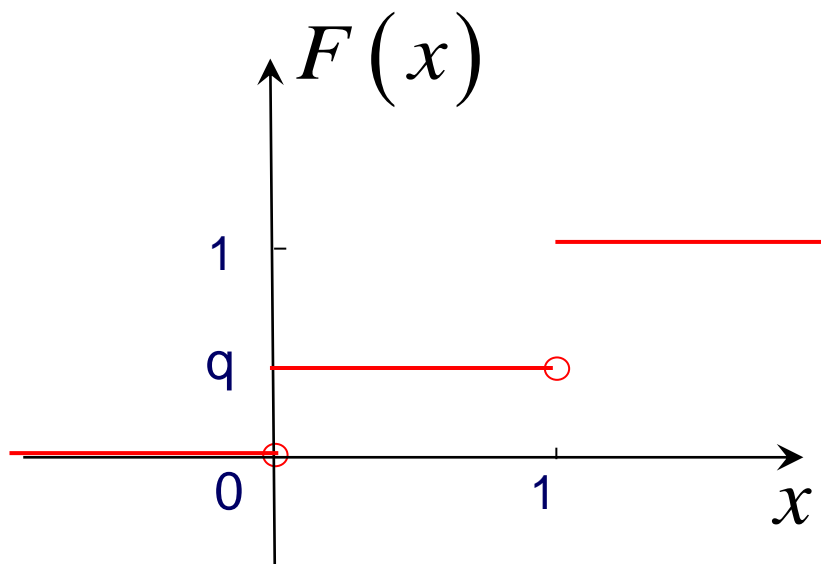
例3. 1 设 $X \sim B(1, p), 0 < p < 1, q = 1 - p,$

X	0	1
p	q	p

求 X 的分布函数 $F(x)$.

解：

$$F(x) = P\{X \leq x\} = \begin{cases} 0 & x < 0 \\ q & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$



一般地，设离散型随机变量 X 的分布律为

$$P\{X = x_k\} = p_k, \quad k = 1, 2, \dots$$

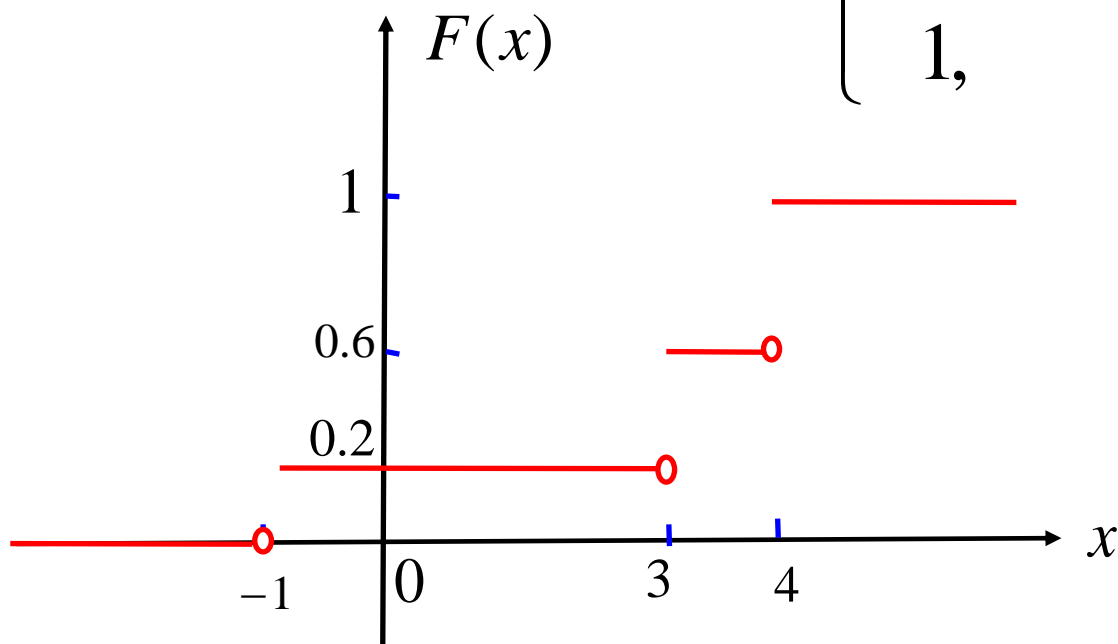
由概率的可列可加性得 X 的分布函数为

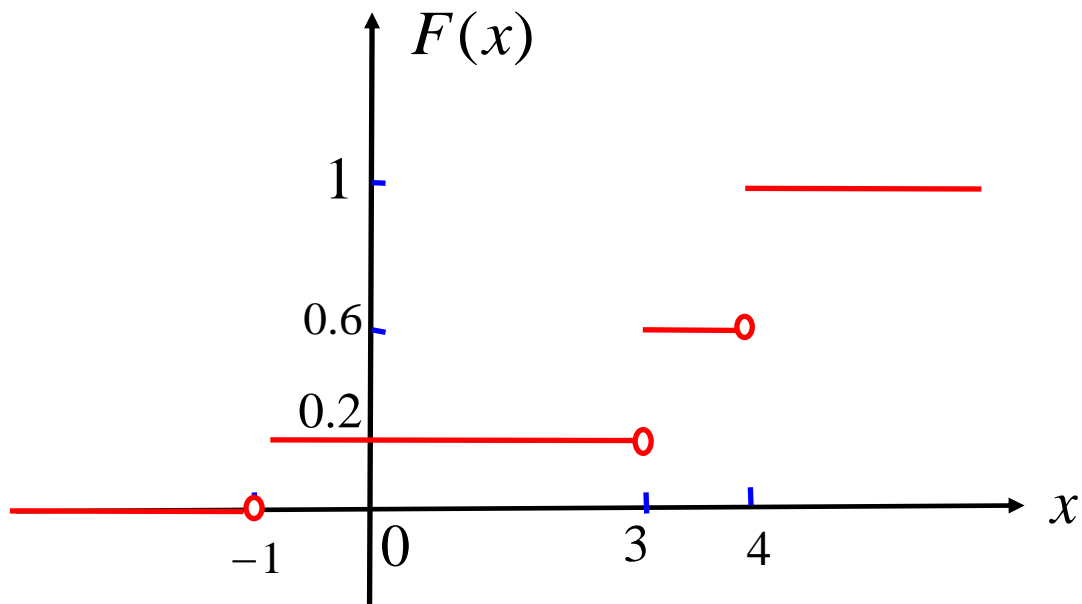
$$F(x) = \sum_{x_k \leq x} p_k$$

分布函数 $F(x)$ 在 $x = x_k, (k = 1, 2, \dots)$ 处有跳跃，其跳跃值为 $p_k = P\{X = x_k\}$.

例3.2 设随机变量 X 的分布函数如下，
求 X 的分布律.

$$F(x) = \begin{cases} 0, & x < -1, \\ 0.2, & -1 \leq x < 3, \\ 0.6, & 3 \leq x < 4, \\ 1, & x \geq 4. \end{cases}$$





解： $F(x)$ 只在 $-1, 3, 4$ 有跳，跳的幅度分别是
0.2, 0.4, 0.4. \therefore 分布律为

X	-1	3	4
p	0.2	0.4	0.4

例3.3 设一物体在 A, B 两点间移动, A, B 之间距离3个单位。该物体落在 A, B 间任一子区间的概率与区间长度成正比。设它离 A 点的距离为 X , 求 X 的分布函数。

解：根据题意， $P(0 \leq X \leq 3) = 1$,

$$P(0 \leq X \leq 3) = 3k = 1, \Rightarrow k = \frac{1}{3},$$

$$\text{当 } x < 0 \text{ 时, } F(x) = P(X \leq x) = 0,$$

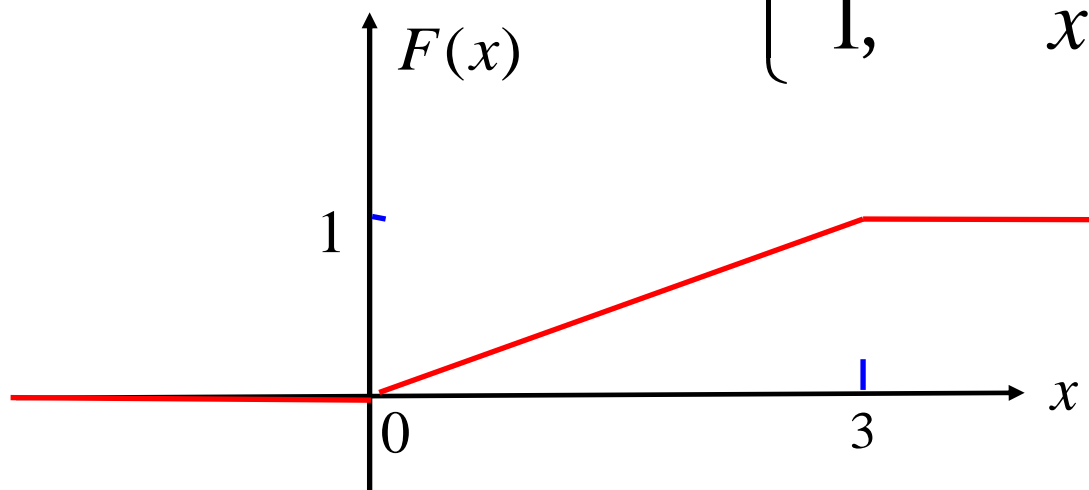
$$\text{当 } x \geq 3 \text{ 时, } F(x) = P(X \leq x) = 1,$$

$$\text{当 } 0 \leq x < 3 \text{ 时, } F(x) = P(X \leq x)$$

$$= P(X < 0) + P(0 \leq X \leq x) = \frac{x}{3}.$$

X 的分布函数为

$$F(x) = P(X \leq x) = \begin{cases} 0, & x < 0 \\ \frac{x}{3}, & 0 \leq x < 3. \\ 1, & x \geq 3 \end{cases}$$



与离散型随机变量的分布函数不同!

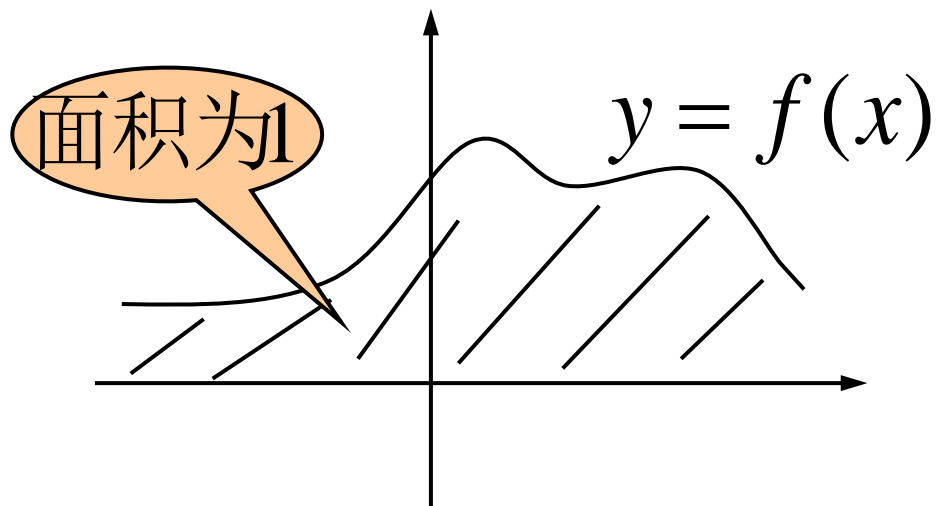
2.4 连续型随机变量及其概率密度函数

定义:对于随机变量 X 的分布函数 $F(x)$,若存在非负的函数 $f(x)$,使对于任意实数 x ,有:

$$F(x) = \int_{-\infty}^x f(t) dt$$

则称 X 为连续型随机变量,其中 $f(x)$ 称为 X 的概率密度函数,简称密度函数.

$f(x)$ 的性质:



1) $f(x) \geq 0$,

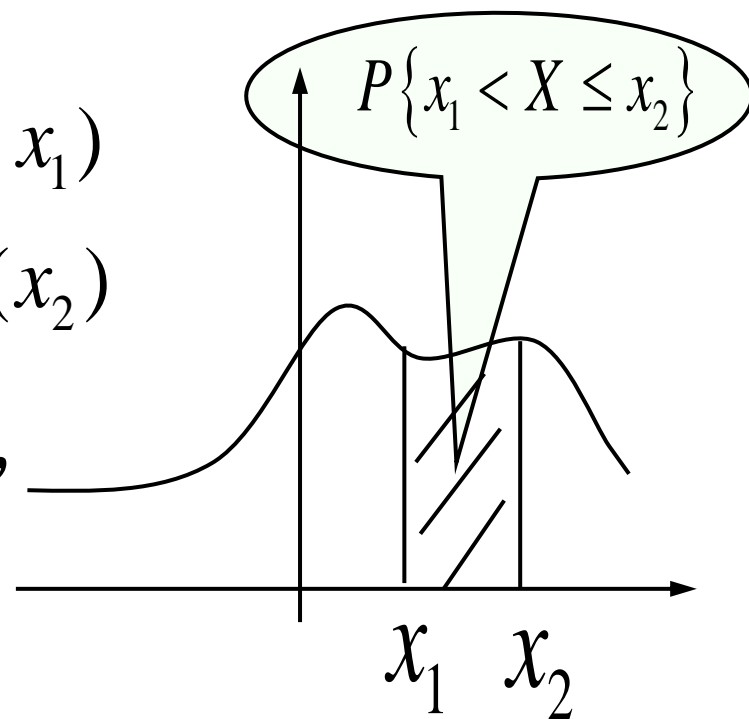
2) $\int_{-\infty}^{+\infty} f(x)dx = 1$,

3) 对于任意的实数 $x_1, x_2 (x_2 > x_1)$

$$P\{x_1 < X \leq x_2\} = F(x_2) - F(x_1)$$

$$= \int_{x_1}^{x_2} f(t) dt,$$

$$\Rightarrow P(X = a) = 0.$$



$f(x)$ 的性质:

4) 在 $f(x)$ 连续点 x , $F'(x) = f(x)$

即在 $f(x)$ 的连续点

$$f(x) = F'(x) = \lim_{\Delta x \rightarrow 0} \frac{F(x + \Delta x) - F(x)}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{P(x < X \leq x + \Delta x)}{\Delta x}$$

与物理学中的质量线密度的定义相类似

$$P(x < X \leq x + \Delta x) \approx f(x) \cdot \Delta x$$

这表示 X 落在点 x 附近 $(x, x + \Delta x]$ 的概率近似等于 $f(x)\Delta x$

思考题：

设 A, B 为随机事件，

若 $P(A) = 1$ ，则 A 为必然事件吗？

若 $P(B) = 0$ ，则 B 为不可能事件吗？

若 $P(AB) = 0$ ，则 A 与 B 不相容吗？

答：都不一定。例如：

$$X \in [0, 1], f(x) = \begin{cases} 1, & x \in [0, 1], \\ 0, & \text{其他.} \end{cases} \quad A = \{0 < X < 1\},$$

$$B = \{X = 0.5\}, \text{ 则 } P(A) = 1, \quad P(B) = 0, \quad P(AB) = 0.$$

例4.1 设 X 的密度函数为 $f(x) = \begin{cases} c, & 0 < x < 1, \\ 2/9, & 3 < x < 6, \\ 0, & \text{其他.} \end{cases}$

(1) 求常数 c 的值;

(2) 写出 X 的分布函数;

(3) 要使 $P(X < k) = \frac{2}{3}$, 求 k 的值。

解： (1) $1 = \int_{-\infty}^{+\infty} f(t)dt$

$$= c \int_0^1 dt + \frac{2}{9} \int_3^6 dt = \frac{2}{3} + c \Rightarrow c = \frac{1}{3}$$

(2) $F(x) = P\{X \leq x\} = \int_{-\infty}^x f(t)dt$

$$= \begin{cases} \int_{-\infty}^x 0dt, & x < 0, \\ \int_{-\infty}^0 0dt + \int_0^x \frac{1}{3} dt, & 0 \leq x < 1, \\ \int_{-\infty}^0 0dt + \int_0^1 \frac{1}{3} dt + \int_1^x 0dt, & 1 \leq x < 3, \\ \int_{-\infty}^0 0dt + \int_0^1 \frac{1}{3} dt + \int_1^3 0dt + \int_3^x \frac{2}{9} dt, & 3 \leq x < 6, \\ \int_{-\infty}^0 0dt + \int_0^1 \frac{1}{3} dt + \int_1^3 0dt + \int_3^6 \frac{2}{9} dt + \int_6^x 0dt, & x \geq 6. \end{cases}$$

$$(2) \quad F(x) = \int_{-\infty}^x f(t) dt = \begin{cases} 0, & x < 0, \\ x/3, & 0 \leq x < 1, \\ 1/3, & 1 \leq x < 3, \\ (2x-3)/9, & 3 \leq x < 6, \\ 1, & x \geq 6. \end{cases}$$

$$(3) \quad \text{使 } P(X < k) = \frac{2}{3}, \quad \text{即 } F(k) = \frac{2}{3},$$

$$\Rightarrow 3 < k < 6, \quad \therefore \frac{2k-3}{9} = \frac{2}{3}, \quad k = 4.5.$$

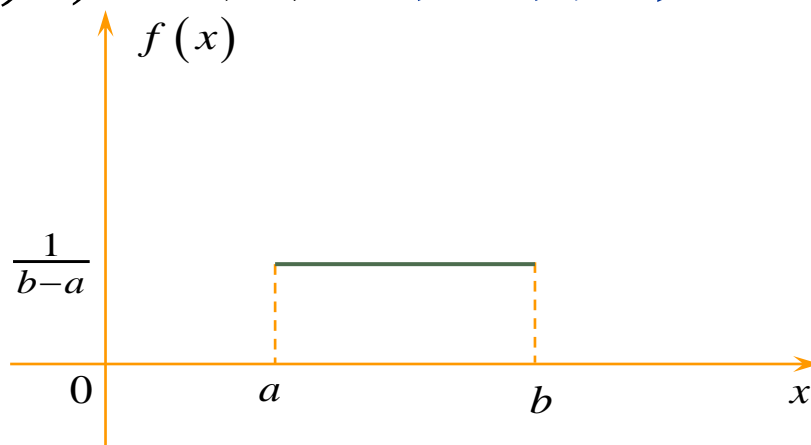
几个重要的连续型随机变量分布

一、均匀分布

定义：设随机变量 X 具有概率密度函数

$$f(x) = \begin{cases} \frac{1}{b-a}, & x \in (a, b), \\ 0, & \text{其他.} \end{cases}$$

称 X 在区间 (a, b) 上服从均匀分布，记为 $X \sim U(a, b)$.

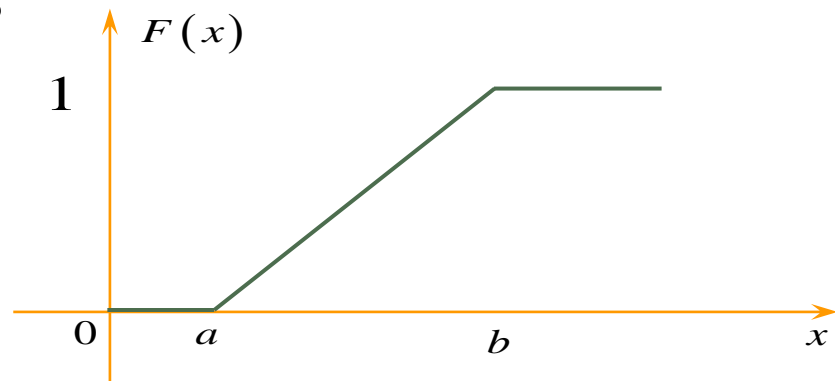


性质： 设 $a \leq c < c+l \leq b$

$$\Rightarrow P(c < X < c+l) = \int_c^{c+l} \frac{1}{b-a} dt = \frac{l}{b-a} \quad \text{---与} c \text{无关}$$

分布函数

$$F(x) = \begin{cases} 0, & x \leq a, \\ \frac{x-a}{b-a}, & a < x < b, \\ 1, & x \geq b. \end{cases}$$



例4.2 (1) 在区间 $(-1, 2)$ 上随机取一数 X ,
试写出 X 的概率密度函数。并求 $P(X > 0)$
的值;

(2) 若在该区间上独立随机取10个数, 求
10个数中恰有两个数大于0的概率。

解：(1) 根据题意， X 在区间 $(-1, 2)$ 上均匀分布

$$\Rightarrow f(x) = \begin{cases} \frac{1}{3}, & -1 < x < 2 \\ 0, & \text{其他} \end{cases} \longrightarrow P(X > 0) = \frac{2}{3},$$

(2) 设10个数中有 Y 个数大于0， 则：

$$Y \sim B(10, \frac{2}{3}) \Rightarrow P(Y = 2) = C_{10}^2 \left(\frac{2}{3}\right)^2 \left(\frac{1}{3}\right)^8$$

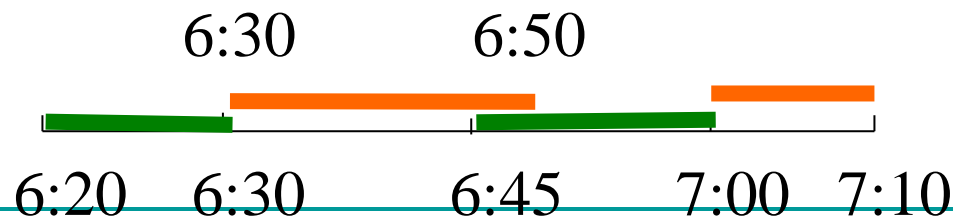
例4.3 杭州某长途汽车站每天从早上6点（第一班车）开始，每隔30分钟有一班车开往上海。

王先生在早上6:20过 X 分钟到达车站，设 X 服从 $(0, 50)$ 上的均匀分布，

(1) 求王先生候车时间不超过15分钟的概率；

(2) 如果王先生一月中有两次按此方式独立地去候车，求他一次候车不超过15分钟，另一次候车大于10分钟的概率。

解： (1) $P(\text{候车时间不超过15钟})=25/50=0.5$,
 (2) $P(\text{候车时间大于10分钟})=30/50=3/5$,
 $P(\text{一次候车时间不超过15分钟, 另一次大于10分钟}) = P((X_1 < 15, X_2 > 10) \cup (X_1 > 10, X_2 < 15))$
 $= P(X_1 < 15, X_2 > 10) + P(X_1 > 10, X_2 < 15)$
 $- P(10 < X_1 < 15, 10 < X_2 < 15)$
 $= 0.5 \times 3/5 + 3/5 \times 0.5 - 0.1 \times 0.1 = 0.59.$



二、正态分布

定义：设 X 的概率密度函数为

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < +\infty$$

其中 $-\infty < \mu < \infty, \sigma > 0$ 为常数，称 X 服从
参数为 μ, σ 的**正态分布(Gauss分布)**，
记为 $X \sim N(\mu, \sigma^2)$

可以验证: $\int_{-\infty}^{+\infty} f(x)dx = 1$

$$\because \int_{-\infty}^{+\infty} f(x)dx \stackrel{\text{令 } t = \frac{x-\mu}{\sigma}}{=} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{t^2}{2}} \sigma dt$$

$$= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

记 $I = \int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} dt$, 只需证明 $I = \sqrt{2\pi}$.

$$\because I^2 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-\frac{(x^2+y^2)}{2}} dx dy$$

$$= \int_0^{2\pi} d\theta \int_0^{+\infty} r e^{-\frac{r^2}{2}} dr = 2\pi \times 1 = 2\pi$$

$$\Rightarrow I = \sqrt{2\pi} \quad \Rightarrow \int_{-\infty}^{+\infty} f(x) dx = 1.$$

性质：若 $X \sim N(\mu, \sigma^2)$

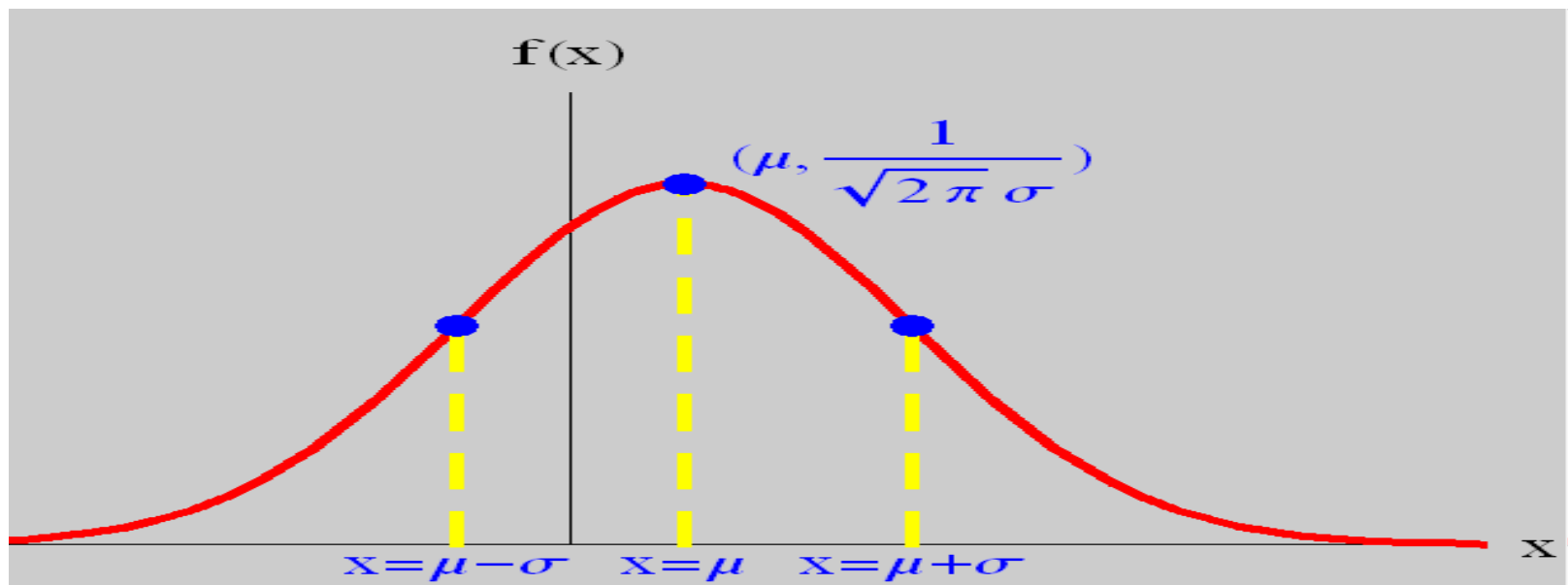
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < +\infty$$

则 1° $f(x)$ 关于 $x = \mu$ 对称

$$2^\circ \quad f_{\max} = f(\mu) = \frac{1}{\sqrt{2\pi}\sigma}$$

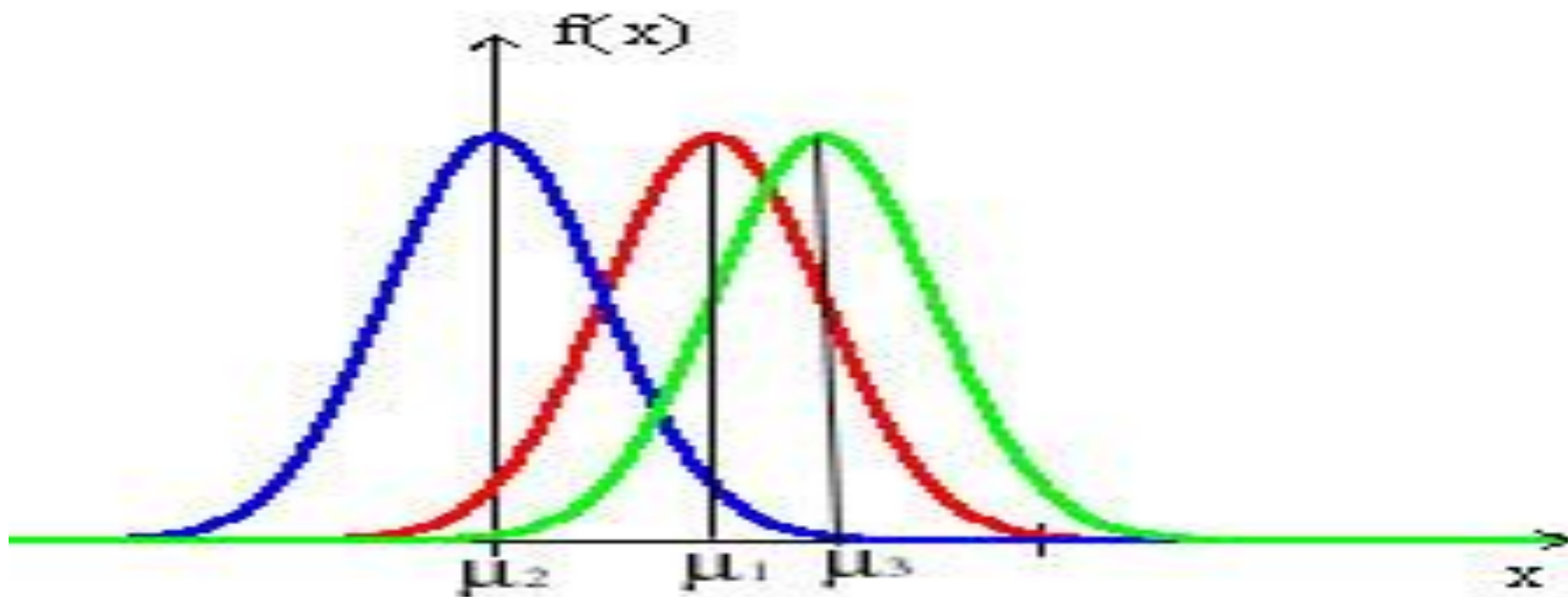
$$3^\circ \quad \lim_{|x-\mu| \rightarrow \infty} f(x) = 0$$

正态概率密度函数

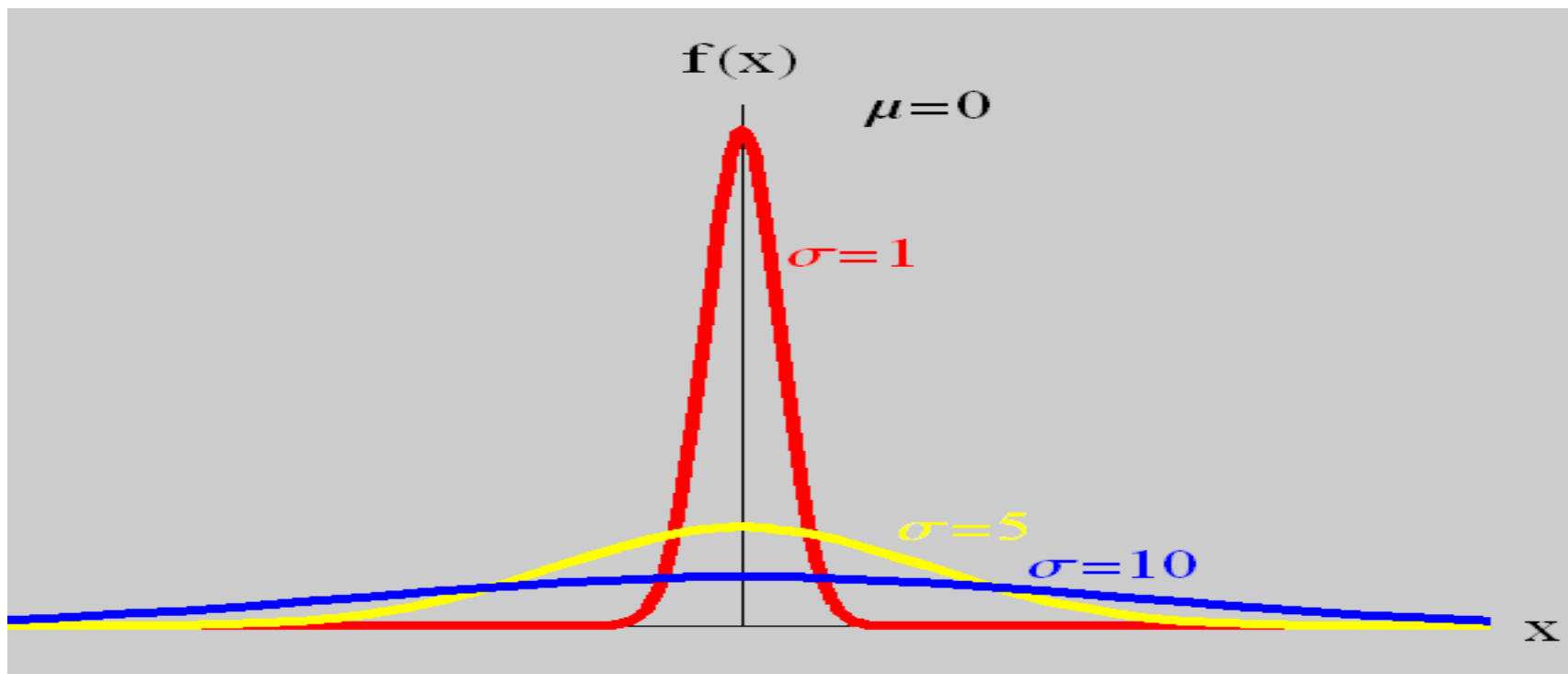


$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < +\infty$$

当固定 σ ，改变 μ 的大小时，密度函数图形形状不变，只是沿着 x 轴作平移变换；



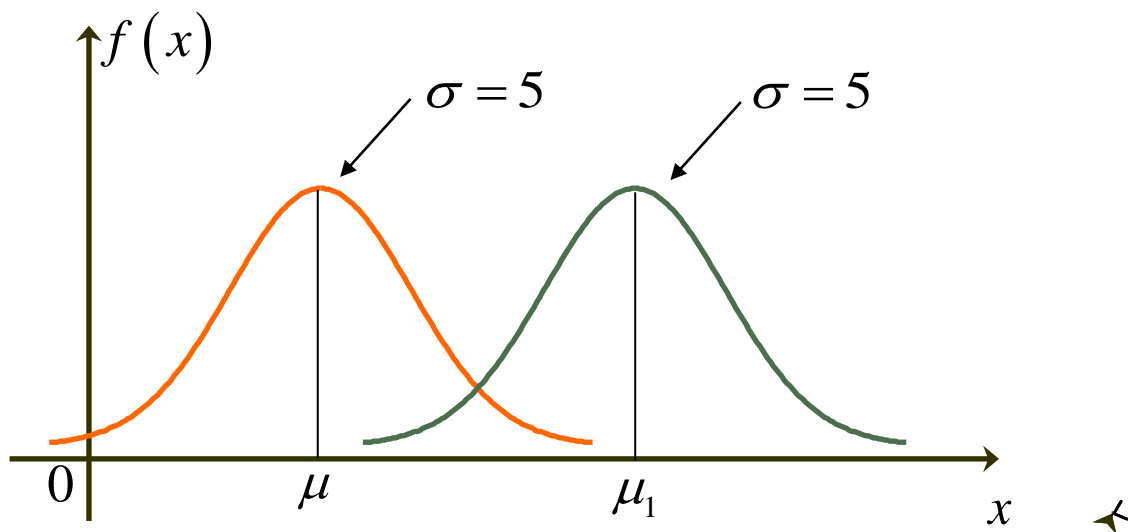
当固定 μ ，改变 σ 的大小时，密度函数图形的对称轴不变，形状发生改变， σ 越小，图形越高瘦， σ 越大，图形越矮胖。



$$X \sim N(\mu, \sigma^2)$$

称 μ 为位置参数(决定对称轴位置)

σ 为尺度参数(决定曲线分散性)



- X 的取值呈中间多，两头少，对称的特性。
- 当固定 μ 时， σ 越大，曲线的峰越低，落在 μ 附近的概率越小，取值就越分散，即 σ 是反映 X 的取值分散性的一个指标。
- 在自然现象和社会现象中，大量随机变量服从或近似服从正态分布。

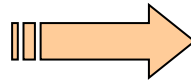
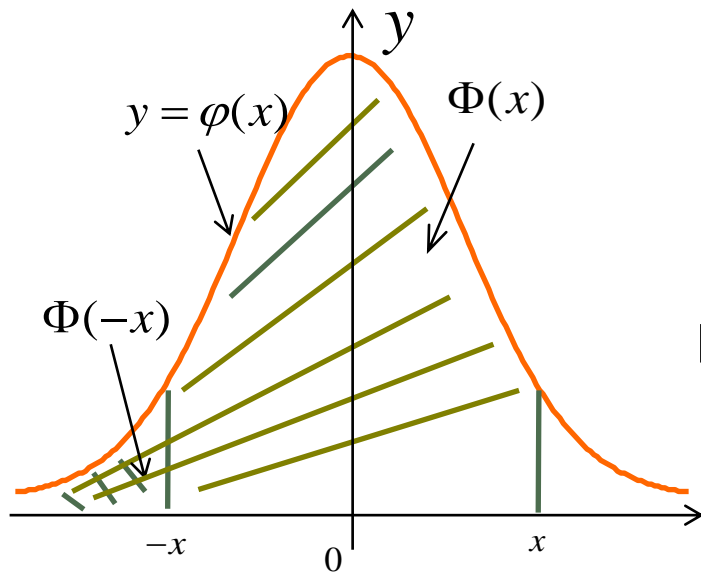
正态分布下的概率计算

$$P\{X \leq x\} = F(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$$
$$= ?$$

✦ 若 $Z \sim N(0, 1)$, 称 Z 服从标准正态分布.

Z 的密度函数为: $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},$

Z 的分布函数为: $\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt.$



$$\Phi(x) + \Phi(-x) = 1$$

✦ 当 $X \sim N(\mu, \sigma^2)$ 时

$$\therefore P(X \leq b) = \int_{-\infty}^b \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

$$\left(\text{作变换: } \frac{x-\mu}{\sigma} = t \right) = \int_{-\infty}^{\frac{b-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

$$\Rightarrow P(X \leq b) = \Phi\left(\frac{b-\mu}{\sigma}\right)$$

当 $X \sim N(\mu, \sigma^2)$ 时 $\frac{X-\mu}{\sigma} \sim N(0, 1)$

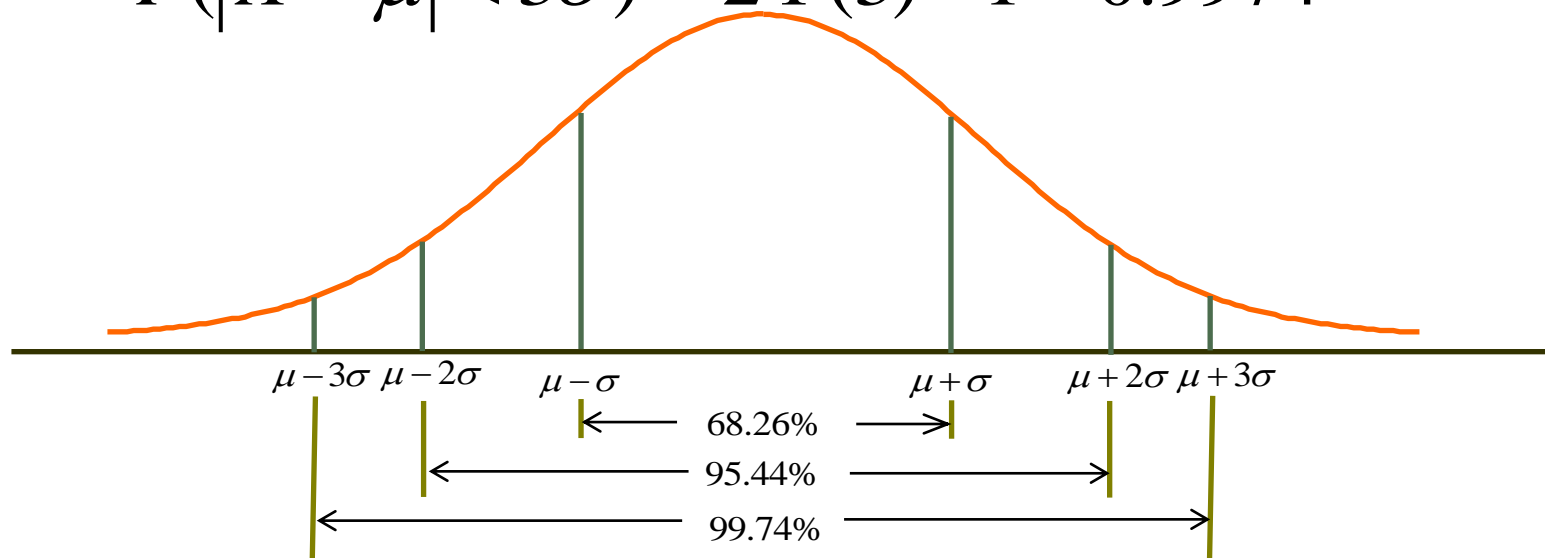
例4.4 $X \sim N(\mu, \sigma^2)$

$$\begin{aligned} P(|X - \mu| < \sigma) &= P(\mu - \sigma < X < \mu + \sigma) \\ &= P\left(\frac{(\mu - \sigma) - \mu}{\sigma} < \frac{X - \mu}{\sigma} < \frac{(\mu + \sigma) - \mu}{\sigma}\right) \end{aligned}$$

$$= \Phi(1) - \Phi(-1) = 2\Phi(1) - 1 = 0.6826$$

$$P(|X - \mu| < 2\sigma) = 2\Phi(2) - 1 = 0.9544$$

$$P(|X - \mu| < 3\sigma) = 2\Phi(3) - 1 = 0.9974$$



例4.5 用天平称一实际重量为 a 的物体，天平的读数为随机变量 X ，设 $X \sim N(a, 0.01^2)$ 时，

(1) 求读数与 a 的误差小于0.005的概率；

(2) 求读数至少比 a 多0.0085的概率。

解:(1) $P(|X - a| < 0.005)$

$$= \Phi\left(\frac{0.005}{0.01}\right) - \Phi\left(-\frac{0.005}{0.01}\right)$$

$$= 2\Phi(0.5) - 1$$

查附表

$$=== 2 \times 0.6915 - 1 = 0.3830$$

(2) $P(X - a \geq 0.0085) = 1 - \Phi(0.85)$

$$= 1 - 0.8023 = 0.1977.$$

注：计算 $P(X - a < 0.0085)$ 使用*Excel*表单：
在*Excel*表单的任一单元格输入

“ $=NORM.DIST(0.0085, 0, 0.01, 1)$ ”

点击“确定”，即在单元格中出现“0.802337508”。

例4.6 一批钢材(线材)长度 $X(\text{cm}) \sim N(\mu, \sigma^2)$

(1) 若 $\mu=100$, $\sigma=2$, 求这批钢材长度小于 97.8cm 的概率;

(2) 若 $\mu=100$, 要使这批钢材的长度至少有 90% 落在区间 (97, 103) 内, 问 σ 至多取何值?

$$\text{解: (1) } P(X < 97.8) = P\left(\frac{X - 100}{2} < \frac{97.8 - 100}{2}\right)$$

$$= \Phi\left(\frac{97.8 - 100}{2}\right) = 1 - \Phi(1.1)$$

查附表

$$= 1 - 0.8643 = 0.1357$$

$$(2) \text{ 需: } P\{97 < X < 103\} \geq 90\%$$

$$\text{即 } \Phi\left(\frac{103 - 100}{\sigma}\right) - \Phi\left(\frac{97 - 100}{\sigma}\right) = 2\Phi\left(\frac{3}{\sigma}\right) - 1 \geq 90\%$$

$$\Rightarrow \Phi\left(\frac{3}{\sigma}\right) \geq 0.95 \Rightarrow \frac{3}{\sigma} \geq 1.645 \Rightarrow \sigma \leq 1.8237$$

例4.7 设一天中经过一高速公路某一入口的重型车辆数 X 近似服从 $N(\mu, \sigma^2)$ ，已知有25%的天数超过400辆，有33%的天数不到350辆，求 μ, σ .

解： 已知 $P(X > 400) = 0.25$, $P(X < 350) = 0.33$

$$\text{而 } P(X > 400) = 1 - \Phi\left(\frac{400 - \mu}{\sigma}\right) = \Phi\left(\frac{\mu - 400}{\sigma}\right),$$

$$P(X < 350) = \Phi\left(\frac{350 - \mu}{\sigma}\right),$$

$$\Phi(-0.675) = 0.25, \quad \Phi(-0.440) = 0.33,$$

$$\text{于是 } \begin{cases} \frac{\mu - 400}{\sigma} = -0.675 \\ \frac{350 - \mu}{\sigma} = -0.440 \end{cases} \Rightarrow \begin{cases} \mu \approx 369.7 \\ \sigma \approx 44.8 \end{cases}$$

三、指数分布

定义：设 X 的密度函数为 $f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0, \\ 0, & x \leq 0. \end{cases}$

其中 $\lambda > 0$ 为常数，则称 X 服从参数为 λ 的指数分布。记为 $X \sim E(\lambda)$ 或 $X \sim \text{Exp}(\lambda)$.

X 的分布函数为：

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x > 0, \\ 0, & x \leq 0. \end{cases}$$

X 具有如下的**无记忆性**： $t_0 > 0, t > 0$,

$$\begin{aligned} P(X > t_0 + t \mid X > t_0) &= \frac{P(X > t_0 + t)}{P(X > t_0)} \\ &= \frac{1 - F(t_0 + t)}{1 - F(t_0)} = e^{-\lambda t} = P(X > t). \end{aligned}$$

如果 X 表示等待时间，那么无记忆性说明只要还没等到，那么剩余等待时间仍然服从参数为 λ 的指数分布。

如果 X 表示元件寿命，那么无记忆性说明只要还没坏掉，那么剩余寿命仍然服从参数为 λ 的指数分布。

例4.8 某大型设备在任何长度为 t 的时间区间内发生故障的次数 $N(t)$ 服从参数为 λt 的泊松分布.记设备无故障运行的时间为 T ,

(1)求 T 的概率分布函数;

(2)已知设备无故障运行了10个小时, 求该设备再无故障至少运行8个小时的概率.

解： (1) $P\{N(t) = k\} = e^{-\lambda t} (\lambda t)^k / k!, k = 0, 1, 2, \dots$

$$F_T(t) = P\{T \leq t\} = 1 - P\{T > t\}$$

当 $t < 0$ 时, $F_T(t) = 0$

当 $t \geq 0$ 时, $F_T(t) = 1 - P\{N(t) = 0\} = 1 - e^{-\lambda t}$

$$(2) P\{T \geq 18 | T > 10\} = \frac{P\{T > 18\}}{P\{T > 10\}} = e^{-8\lambda} = P\{T > 8\}$$

例4.9 一银行服务需要等待，设等待时间 X (分钟)的概率密度函数为

$$f(x) = \begin{cases} \frac{1}{10} e^{\frac{-x}{10}}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

某人进了银行，且打算过会儿去办另一件事，于是先等待，如果超过15分钟还没有等到服务就离开，设他实际的等待时间为 Y ，(1)求 Y 的分布函数;(2)问 Y 是离散型随机变量吗？连续型随机变量吗？

解: (1) $0 \leq Y \leq 15$, $F(y) = P(Y \leq y)$

当 $y < 0$ 时, $F(y) = P(Y \leq y) = 0$,

当 $y \geq 15$ 时, $F(y) = P(Y \leq y) = 1$,

当 $0 \leq y < 15$ 时, $F(y) = P(Y \leq y)$
$$= P(X \leq y) = 1 - e^{\frac{-y}{10}}.$$

(2) Y 的取值范围为 $[0, 15]$, 故不是离散型随机变量;

$$\begin{aligned} P\{Y = 15\} &= P(Y \leq 15) - P(Y < 15) \\ &= F(15) - F(15 - 0) = e^{-1.5} > 0, \end{aligned}$$

因此 Y 也不是连续型随机变量; Y 既没有分布律, 也没有概率密度函数.

2.5 随机变量函数的分布

例如，若要测量一个圆的面积 Y ，可以先测量其周长，如果周长的测量值 X 是随机变量，有一个概率分布，那么 Y 服从什么分布？

问题：已知随机变量 X 的概率分布，

且已知 $Y=g(X)$ ，求 Y 的概率分布。

例5.1 已知 X 具有分布律

X	-1	0	1
p	0.2	0.5	0.3

且设 $Y=X^2$ ，求 Y 的概率分布。

解： Y 的所有可能取值为0, 1

$$P(Y = 0) = P(X = 0) = 0.5$$

$$\begin{aligned} P(Y = 1) &= P\{(X = 1) \cup (X = -1)\} \\ &= P(X = 1) + P(X = -1) = 0.5 \end{aligned}$$

即找出 $(Y=0)$ 的等价事件 $(X=0)$ ；

$(Y=1)$ 的等价事件 $(X=1)$ 与 $(X=-1)$ 的和事件.

例5.2 设随机变量 X 具有概率密度函数

$$f_X(x) = \begin{cases} \frac{x}{8}, & 0 < x < 4, \\ 0, & \text{其他.} \end{cases}$$

求 $Y = X^2$ 的概率密度函数。

解：记 Y 的分布函数为 $F_Y(y)$

$$F_Y(y) = P\{Y \leq y\} = P\{X^2 \leq y\}$$

当 $y \leq 0$ 时, $F_Y(y) = 0$;

当 $y \geq 16$ 时, $F_Y(y) = 1$

当 $0 < y < 16$ 时,

$$\begin{aligned} F_Y(y) &= P\{X^2 \leq y\} = P\{-\sqrt{y} \leq X \leq \sqrt{y}\} \\ &= P\{-\sqrt{y} \leq X \leq 0\} + P\{0 < X \leq \sqrt{y}\} \\ &= P\{0 < X \leq \sqrt{y}\} = \int_0^{\sqrt{y}} \frac{t}{8} dt = \frac{y}{16} \\ \Rightarrow f_Y(y) &= \begin{cases} \frac{1}{16}, & 0 < y < 16 \\ 0, & \text{其他} \end{cases} \end{aligned}$$

Y 在区间 $(0, 16)$ 上均匀分布。

一般，若已知 X 的概率分布， $Y=g(X)$ ，求 Y 的概率分布的过程为：

1. 若 Y 为离散型随机变量，则先写出 Y 的可能取值： $y_1, y_2, \cdots y_j, \cdots$ ，再找出 $(Y = y_j)$ 的等价事件 $(X \in D_j)$ ，得 $P(Y = y_i) = P(X \in D_j)$ ；
2. 若 Y 为连续型随机变量，则先写出 Y 的概率分布函数： $F_Y(y) = P(Y \leq y)$ ，找出 $(Y \leq y)$ 的等价事件 $(X \in D_y)$ ，得 $F_Y(y) = P(X \in D_y)$ ；再求出 Y 的概率密度函数 $f_Y(y)$ 。

关键是找出等价事件。

例5.3 设随机变量 X 的分布律如下表

X	-1	0	1	2
p	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{3}$

$Y=2X+1$, $Z=X^2$, 求 Y, Z 的概率分布律.

解： Y 的可能取值为 $-1, 1, 3, 5$,

Z 的可能取值为 $0, 1, 4$,

$(Y=-1)$ 的等价事件为 $(X=-1) \cdots$

$(Z=1)$ 的等价事件为 $(X=1) \cup (X=-1)$

故得：

Y	-1	1	3	5
p	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{3}$

Z	0	1	4
p	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{3}$

例5.4: 设随机变量 X 具有密度函数

$$f_X(x), -\infty < x < +\infty,$$

分别求 $Y = |X|$, $Z = X^2$ 的

概率密度函数 $f_Y(y)$, $f_Z(z)$.

解： 分别记 X, Y, Z 的分布函数为 $F_X(x), F_Y(y), F_Z(z)$.

当 $y \leq 0$ 时, $F_Y(y) = 0$. 当 $y > 0$ 时,

$$F_Y(y) = P\{Y \leq y\} = P\{|X| \leq y\} = F_X(y) - F_X(-y).$$

$$f_Y(y) = \begin{cases} f_X(y) + f_X(-y), & y > 0 \\ 0, & y \leq 0 \end{cases}$$

同理，当 $z \leq 0$ 时， $F_Z(z) = 0$ 。当 $z > 0$ 时，

$$F_Z(z) = P\{Z \leq z\} = P\{X^2 \leq z\}$$

$$= P\{-\sqrt{z} \leq X \leq \sqrt{z}\} = F_X(\sqrt{z}) - F_X(-\sqrt{z})$$

$$f_Z(z) = \begin{cases} \frac{1}{2\sqrt{z}}[f_X(\sqrt{z}) + f_X(-\sqrt{z})], & z > 0 \\ 0 & , \quad z \leq 0. \end{cases}$$

例5.5 设 $X \sim U(-1, 2)$, 求

$Y = |X|$ 的概率密度函数 $f_Y(y)$.

解：X的密度函数为 $f_X(x) = \begin{cases} 1/3, & -1 < x < 2 \\ 0, & \text{其它} \end{cases}$

由例5.4得 $Y = |X|$ 的概率密度函数为

$$f_Y(y) = \begin{cases} f_X(y) + f_X(-y), & y > 0 \\ 0, & y \leq 0 \end{cases}$$
$$= \begin{cases} \frac{1}{3} + \frac{1}{3}, & 0 < y < 1 \\ \frac{1}{3} + 0, & 1 < y < 2 \\ 0, & \text{其它} \end{cases} = \begin{cases} \frac{2}{3}, & 0 < y < 1 \\ \frac{1}{3}, & 1 < y < 2 \\ 0, & \text{其它} \end{cases}$$

例5.6 设 $X \sim N(0, 1)$, 求
 $Z = X^2$ 的概率密度函数 $f_Z(z)$.

解： X 的密度函数为 $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, -\infty < x < +\infty,$

由例5.4得 $Z = X^2$ 的概率密度函数为

$$f_Z(z) = \begin{cases} \frac{1}{2\sqrt{z}} [f_X(\sqrt{z}) + f_X(-\sqrt{z})], & z > 0 \\ 0 & , \quad z \leq 0. \end{cases}$$

$$= \begin{cases} \frac{1}{\sqrt{2\pi}} z^{-\frac{1}{2}} e^{-\frac{z}{2}}, & z > 0 \\ 0 & , \quad z \leq 0. \end{cases}$$

此时称 Z 服从自由度为 1 的 χ^2 分布。

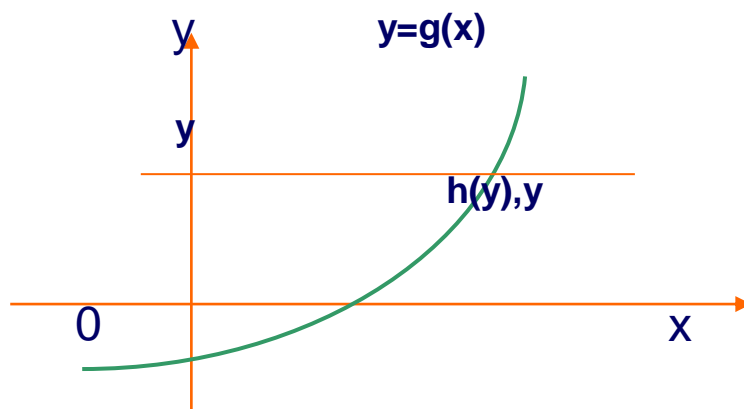
定理2.5.1: 设 $X \sim f_X(x)$, $-\infty < x < +\infty$, $g'(x) > 0$ (或 $g'(x) < 0$)。

$Y = g(X)$, 则 Y 具有概率密度函数为:

$$f_Y(y) = \begin{cases} f_X(h(y)) \cdot |h'(y)|, & \alpha < y < \beta \\ 0, & \text{其他} \end{cases}$$

这里 (α, β) 是 Y 的取值范围, h 是 g 的反函数

$$h(y) = x \Leftrightarrow y = g(x)$$



证明：不妨设 $g'(x) > 0$, 则 $g(x)$ 为单调增函数,

且: $h'(y) > 0$

当 $y \leq \alpha$ 时, $F_Y(y) = 0$; 当 $y \geq \beta$ 时, $F_Y(y) = 1$;

当 $\alpha < y < \beta$ 时,

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y)$$

$$= P(X \leq h(y)) = F_X(h(y))$$

$$\Rightarrow f_Y(y) = f_X(h(y))h'(y) = f_X(h(y)) \cdot |h'(y)|$$

同理可证：当 $g'(x) < 0$ 时, 定理为真

例5.7 设 $X \sim N(\mu, \sigma^2)$, $Y = aX + b (a \neq 0)$,

求 Y 的概率密度函数 $f_Y(y)$.

解: $y = g(x) = ax + b, \quad g'(x) = a \neq 0,$

$$x = h(y) = \frac{y - b}{a}$$

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y - b}{a}\right)$$

$$= \frac{1}{|a|} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{\left(\frac{y - b}{a} - \mu\right)^2}{2\sigma^2}\right\}$$

$$= \frac{1}{\sqrt{2\pi} |a| \sigma} e^{-\frac{[y - (a\mu + b)]^2}{2a^2\sigma^2}}$$

$$\Rightarrow Y \sim N(a\mu + b, a^2\sigma^2)$$

一般若 $X \sim N(\mu, \sigma^2)$,

$$Y = aX + b \Rightarrow Y \sim N(a\mu + b, a^2\sigma^2)$$

例5.8 设 $X \sim U(-\pi / 2, \pi / 2), Y = \sin X,$

求 Y 的概率密度函数 $f_Y(y)$.

解： $Y = \sin X$ 对应的函数 $y = g(x)$ 在 $(-\pi/2, \pi/2)$

上恒有 $g'(x) = \cos x > 0$, 且有反函数

$$x = h(y) = \arcsin y, h'(y) = 1 / \sqrt{1 - y^2}$$

X 的概率密度函数为

$$f_X(x) = \begin{cases} \frac{1}{\pi}, & -\frac{\pi}{2} < x < \frac{\pi}{2} \\ 0, & \text{其它} \end{cases}$$

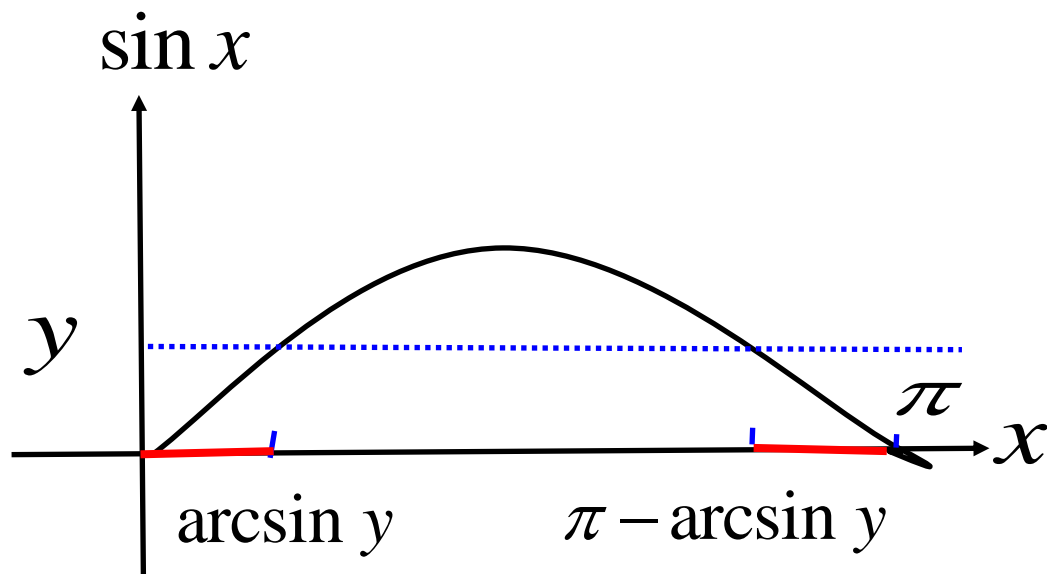
由定理得 $Y = \sin X$ 的密度函数为

$$f_Y(y) = \begin{cases} \frac{1}{\pi} \cdot \frac{1}{\sqrt{1-y^2}}, & -1 < y < 1 \\ 0, & \text{其它} \end{cases}$$

例5.9 $Y = \sin X, X \sim U(0, \pi)$, 求 $f_Y(y)$.

解: $Y = \sin X$ 在 $(0, \pi)$ 不单调, 所以不能应用定理。对 $0 < y \leq 1$,

$$F_Y(y) = P(Y \leq y) = P\{\sin X \leq y\} = \frac{2 \arcsin y}{\pi}$$



所以 Y 的概率密度函数为:

$$f_Y(y) = \begin{cases} \frac{2}{\pi\sqrt{1-y^2}}, & 0 < y < 1, \\ 0, & \text{其他.} \end{cases}$$

例5.10 设 $X \sim E(\lambda)$, $F(x)$ 为 X 的分布函数,
记 $Y = F(X)$, 试证 $Y \sim U(0,1)$ (即均匀分布).

解：由 $X \sim E(\lambda), \therefore$ 分布函数 $F(x) = \begin{cases} 1 - e^{-\lambda x}, & x > 0, \\ 0, & x \leq 0. \end{cases}$

当 $x > 0$ 时, $y = F(x) = 1 - e^{-\lambda x} \in (0, 1)$ 单调增,

反函数 $x = F^{-1}(y) = -\frac{1}{\lambda} \ln(1 - y),$

$$\frac{d}{dy} F^{-1}(y) = \frac{1}{\lambda(1 - y)},$$

由定理2.5.1, 当 $0 < y < 1$ 时,

$$f_Y(y) = \lambda \exp\left\{-\lambda\left[-\frac{1}{\lambda} \ln(1 - y)\right]\right\} \frac{1}{\lambda(1 - y)} = 1,$$

$\therefore Y \sim U(0, 1)$. 更一般的结果见书中例2.5.6.

课件待续!

THE
END